

Scalable Communication and Data Persistence Layer for NVM-based Storage Systems

Hiroki Ohtsuji¹, Takuya Okamoto², Erika Hayashi¹, Eiji Yoshida¹

¹Fujitsu Laboratories Ltd.

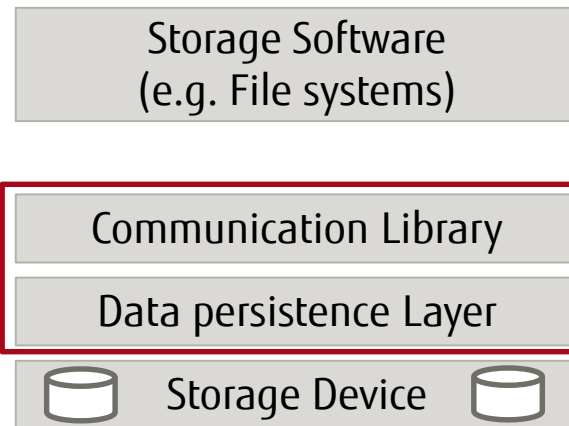
²Fujitsu Ltd.

■ NVM devices

- Short Latency
- Mapped to a memory space
 - No system call for data persistence -> Low overhead

■ Storage Software

- Implemented on the communication and data persistence layers
 - -> The current layer is designed for slow conventional I/O model.

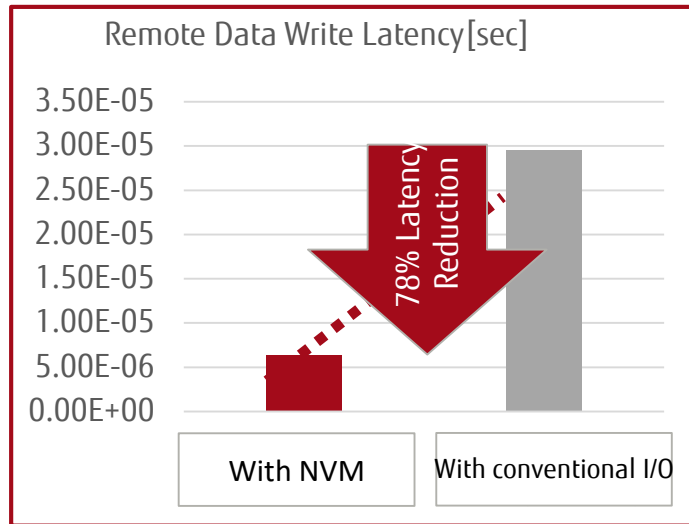
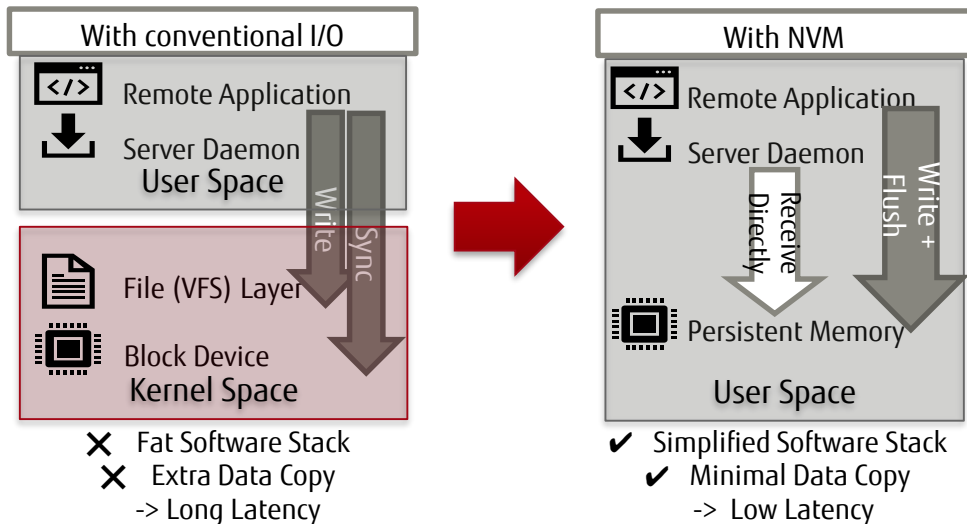


Low-overhead communication and data persistence layer is necessary for storage software with NVM devices

Optimized Remote Data Persistence for NVM devices

Low-overhead operations for remote NVM devices [3]

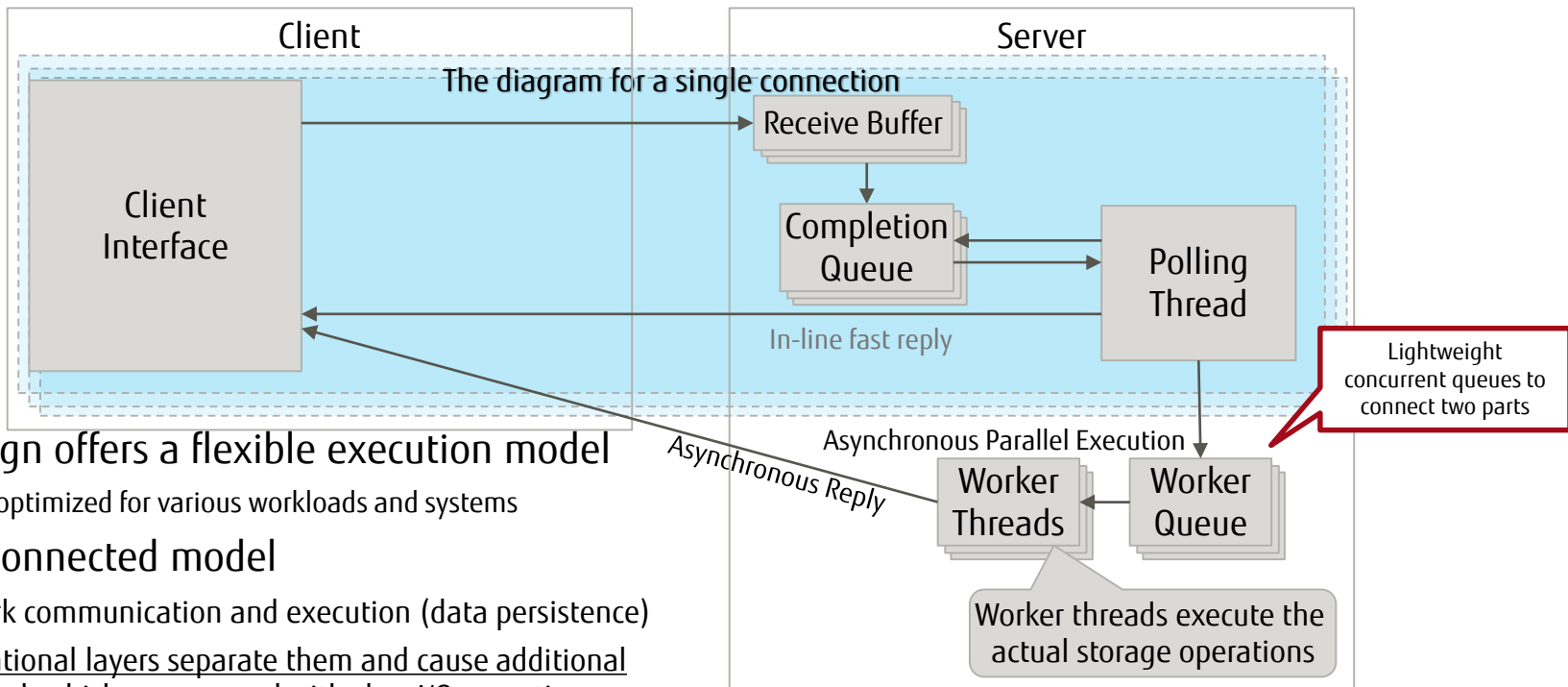
Evaluated with a single client-server connection



Need to consider the scalability of the new persistence model

Scalable Communication and Data Persistence Layer

- We are developing an asynchronous layer for communication and data persistence operations.



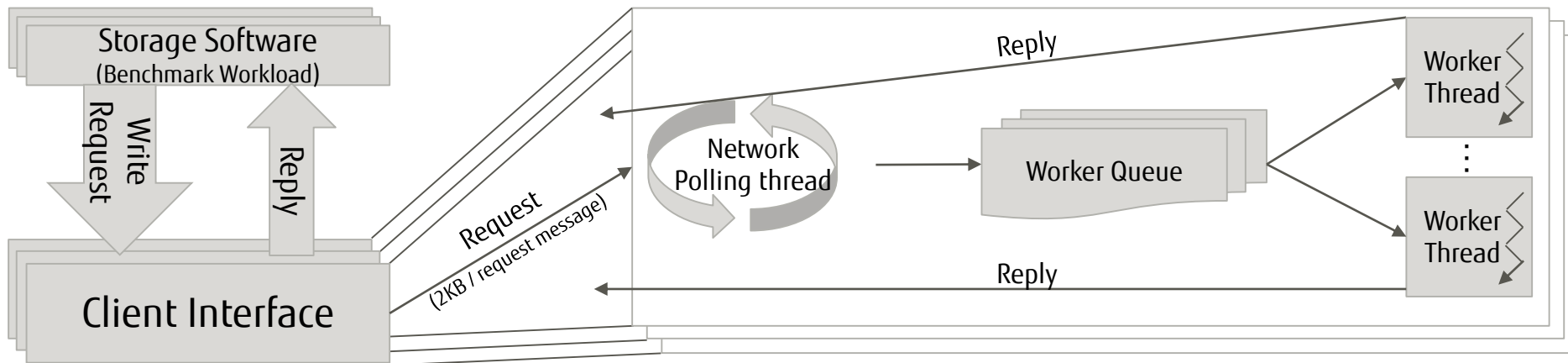
- The design offers a flexible execution model
 - Can be optimized for various workloads and systems
- Tightly connected model
 - Network communication and execution (data persistence)
 - Conventional layers separate them and cause additional overhead, which was covered with slow I/O operations

Using a solid design to minimize the overhead of an asynchronous execution model

Preliminary Evaluation - Environment

- Evaluating the performance of remote write operations
 - Testing the scalability with multiple client processes

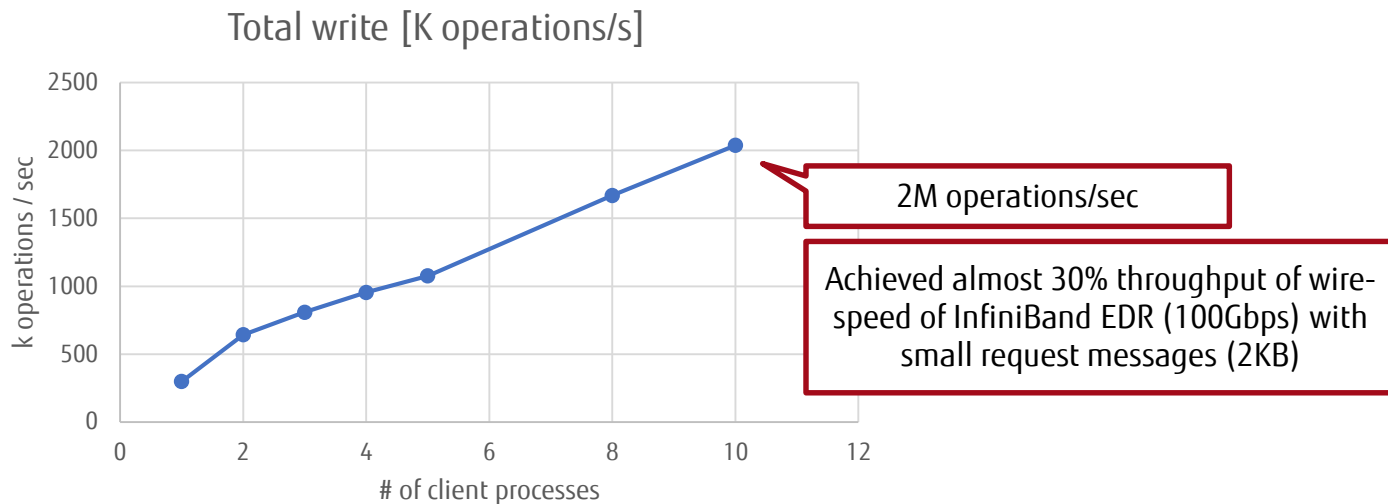
Server: FUJITSU Server PRIMERGY RX2540 M5
CPU: Intel(R) Xeon(R) Gold 6240M CPU @ 2.60GHz x2
NVM: Intel DC Persistent Memory 128 GB x12 *
DRAM: DDR4 DRAM 32 GB x12
Network: Mellanox Connect-X5 EDR HCA Dual Ports



*The data structure for the write benchmark is placed on DRAMs to evaluate the maximum performance of the network and execution layer.

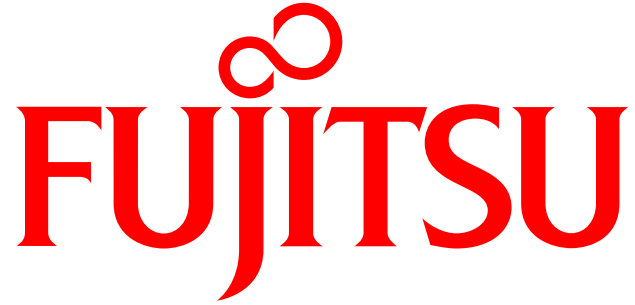
Preliminary Evaluation - Result

- Throughput scalability test with multiple client processes
 - The server processes are running on a single server



Achieving 2M operations /sec with good scalability

- NVM devices require the low-overhead network and execution layer to exploit its performance
- The proposed design tightly connects the network communication layer to the asynchronous execution model for lower overhead
- The preliminary evaluation shows that the architecture shows good scalability and achieved 2M operations / second.
- Future work
 - Detailed evaluation to build the performance model
 - Considering the dynamic method to find the best parameter (number of threads) to exploit the potential of the design

The logo features a red infinity symbol positioned above the word "FUJITSU". The word is rendered in a bold, red, serif typeface. The letter 'J' is particularly stylized, with a long, sweeping tail that extends downwards and to the left.

FUJITSU

shaping tomorrow with you